



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Linear Algebra and its Applications 394 (2005) 73–89

LINEAR ALGEBRA
AND ITS
APPLICATIONS

www.elsevier.com/locate/laa

Permanental bounds for nonnegative matrices via decomposition

George W. Soules

IDA Center for Communications Research, 4320 Westerra Ct, San Diego, CA 92121, USA

Received 25 May 2000; accepted 23 June 2004

Submitted by H. Schneider

Abstract

We investigate an old suggestion of A.E. Brouwer we call *decomposition*, for constructing a class of permanental upper bounds for nonnegative matrices A from a single permanental upper bound $u(B)$ for $(0, 1)$ -matrices B . For certain *feasible* u , which include the Minc–Brègman bound $u(B) = M(B)$ and the Jurkat–Ryser bound $u(B) = J(B)$, we can identify the best and worst of these decomposition bounds. The best decomposition bound, the *star* bound $U^*(A)$, is the only decomposition bound which agrees with u on the $(0, 1)$ -matrices.

If $u = J$, then $U^*(A)$ turns out to be the very bound $U^J(A)$ used by Jurkat and Ryser to obtain J as a special case. If $u = M$, then $U^*(A)$ is a new upper bound $U^M(A)$. We believe its *sharpened* version $U_{\text{6}}^M(A)$ to be the best extant permanental upper bound for nonnegative matrices as well as for $(0, 1)$ -matrices.

© 2004 Elsevier Inc. All rights reserved.

Keywords: Nonnegative matrix; Permanent; Upper bound; Decomposition; Scale symmetries; Sharpening

1. Introduction

We continue our study of the relationship between permanental bounds on \mathcal{B} , the $(0, 1)$ -matrices, and permanental bounds on \mathcal{A} , the nonnegative matrices. We focus on upper bounds, yet with minor modifications these methods extend to lower bounds as well as to functions other than the permanent.

E-mail address: gws@ccrwest.org

0024-3795/\$ - see front matter © 2004 Elsevier Inc. All rights reserved.

doi:10.1016/j.laa.2004.06.022

1.1. Motivation and summary of results

The *permanent* of an n by n matrix $A = [a_{i,j}]$ is $\text{per}(A) = \sum_{\sigma} \prod_{i=1}^n a_{i,\sigma(i)}$ where the sum is over all permutations σ of the full symmetric group. For $A \in \mathcal{A}$ let $A_i = (a_{i,1}, a_{i,2}, \dots, a_{i,n})$ be the i th row of A and $|A_i|$ be the i th row sum $a_{i,1} + \dots + a_{i,n}$; we sometimes let (A_i) denote A . For $k = 1, 2, \dots$ we set $\gamma(k) := (k!)^{1/k}$ with $\gamma(0) = 0$. The Minc–Brègman [1,8] bound

$$M(B) := \prod_{i=1}^n \gamma(|B_i|) \geq \text{per}(B), \quad B \in \mathcal{B} \quad (1)$$

has proven to be a tractable, accurate¹ permanental upper bound on \mathcal{B} , but no comparable bound on \mathcal{A} had appeared until recently [14,15]. There, using induction, concavity, extrapolation and interpolation, we obtained new upper bounds U^1, U^2, U^3, U^4 which were the first to pass the “litmus test” of being tractable and agreeing with M on \mathcal{B} .

Here we investigate a wholly different approach we call *decomposition*, which has already been described briefly in [13] and attributed to A.E. Brouwer. This simple (and, in retrospect, effective) technique produces a class $\{U^\theta(A)\}$ of permanental bounds on \mathcal{A} from a single, prescribed bound $u(B)$ on \mathcal{B} , as follows. Write each row of a given matrix $A \in \mathcal{A}$ as a nonnegative linear combination of $(0, 1)$ -vectors (which decomposition of A we denote by θ). Then apply the row-multilinearity of the permanent to θ to express $\text{per}(A)$ as a nonnegative linear combination of permanents $\{\text{per}(B)\}$ of matrices $B \in \mathcal{B}$. Finally, replace each $\text{per}(B)$ with its upper bound $u(B)$, which results in an expression $U^\theta(A)$, see (26), which perforce bounds $\text{per}(A)$. We call $U^\theta(A)$ the *decomposition bound based on u and θ* . The precise details of this process appear in the proofs.

When the given bound $u(B)$ is *feasible* (Section 2), we can identify the *star* decomposition $\theta = *$, see (5), and the *trivial* decomposition $\theta = 0$, see (4), as yielding the best and worst decomposition bounds $U^*(A)$ and $U^0(A)$ in (27) and (25) respectively, called the *star bound* and the *trivial bound*. The other major results are as follows.

- Under nominal conditions, the star bound $U^*(A)$ is the only decomposition bound agreeing with u on \mathcal{B} . Then there is a 1 to 1 correspondence, $u \leftrightarrow U^*$, between a feasible bound u on \mathcal{B} and its star bound U^* on \mathcal{A} ; U^* is obtained from u via the star decomposition, and u is obtained from U^* by restricting the domain from \mathcal{A} to \mathcal{B} . A corollary is that on \mathcal{B} , no decomposition bound based on u is better than u .
- The Minc–Brègman upper bound $M(B)$ in (1) is feasible, and if $u = M$ we find that $U^*(A)$ is a new permanental upper bound $U^M(A)$, see (6), which is uniformly

¹ Relative to other upper bounds.

better than our previously best (unsharpened) bound $U^4(A)$ given in (19). Our newest, and to date very best permanent upper bound on \mathcal{A} is $U_6^M(A)$, obtained by sharpening $U^M(A)$ (see Section 1.4.1, and the examples in Section 3).

- The Jurkat–Ryser upper bound $J(B)$ in (13) is feasible, and if $u = J$ we find that $U^*(A)$ is in fact the Jurkat–Ryser upper bound $U^J(A)$ in (12), from which J was originally derived by restricting the domain of U^J from \mathcal{A} to \mathcal{B} . In the sense of the 1 to 1 correspondence $J \leftrightarrow U^J$, the Jurkat–Ryser bounds on \mathcal{A} and on \mathcal{B} are equivalent. (The same is true for their lower bounds in (12) and (13).)
- The trivial bound $U^0(A)$ turns out to be a familiar and very weak bound, the product of the row sums of A (the same for each u). It happens that $U^0(B)$ is itself feasible, and for $u = U^0$ we find that every decomposition bound $U^\theta(A)$ based on $U^0(B)$ is the trivial bound $U^0(A)$.

1.2. Generic decompositions

A *decomposition* of a nonnegative row vector $r = (r_j)$ is any way of expressing r as a nonnegative linear combination of $(0, 1)$ -vectors $b = (b_j)$

$$r = \sum_b \theta_b b, \quad (2)$$

from which it follows that $r_j = \sum_b \theta_b b_j$ holds for $j = 1, \dots, n$. A few remarks on this notation are in order. First, the vectors $b \in \{b\}$ are *unindexed*, as we feel indexed notation would be overly cumbersome. And we may assume (by gathering coefficients if need be) that the collection $\{b\}$ has distinct elements, and (by eliminating zero summands if need be) that no $b \in \{b\}$ is zero and that each θ_b is positive. Consequently, the number of summands in (2) will be 0 if $r = 0$, else be between 1 and $2^n - 1$ if $r \neq 0$. Two such decompositions θ, θ' are equal, $\theta = \theta'$, if $\theta_b = \theta'_b$ holds for every b .

A generic row decomposition θ of a matrix $A \in \mathcal{A}$ consists of one such decomposition for each row A_i of A

$$A_i = \sum_{b^i} \theta_{i,b^i} b^i, \quad i = 1, 2, \dots \quad (3)$$

Note that i indexes the rows, not the summands; the coefficient θ , of a $(0, 1)$ -vector $b = b^i$ used in the decomposition of row i , is doubly indexed to indicate it depends on both i and b . Unless otherwise specified, the row index i and column indices j, k, ℓ will range from 1 to n throughout.

1.3. The decompositions $\theta = 0$ and $\theta = *$

Before stating our most general results for bounds based on feasible u and generic decompositions, to fix ideas we compute two bounds on \mathcal{A} based on the

Minc–Brègman bound $u(B) = M(B)$ for the specific decompositions $\theta = 0$ and $\theta = *$, which definitions follow.

The decomposition $\theta = 0$ consists of writing each row A_i of A as the obvious combination of unit (row) vectors $\{e(k) \in \mathcal{B} : |e(k)| = 1\}$

$$A_i = \sum_{k=1}^n a_{i,k} e(k), \quad i = 1, 2, \dots \quad (4)$$

We call $\theta = 0$ the *trivial* decomposition.

For $A \in \mathcal{A}$ let $A^* = [a_{i,j}^*]$ denote the matrix A with each row rearranged into nonincreasing order, $a_{i,1}^* \geq a_{i,2}^* \geq \dots \geq a_{i,n}^* \geq 0$; and for each i let $\pi = \pi^i$ be any permutation of $1, \dots, n$ which so reorders row i , namely $a_{i,\pi(j)} = a_{i,j}^*$ holds for $1 \leq j \leq n$. From (4) we continue, with $a_{i,\pi(n+1)} := 0$,

$$A_i = \sum_{k=1}^n a_{i,\pi(k)} e(\pi(k)) = \sum_{k=1}^n (a_{i,\pi(k)} - a_{i,\pi(k+1)}) \sum_{j=1}^k e(\pi(j)).$$

As $a_{i,\pi(k)} \geq a_{i,\pi(k+1)}$ holds for each k , this gives another decomposition, $\theta = *$, which we write in terms of A^* as

$$A_i = \sum_{k=1}^n (a_{i,k}^* - a_{i,k+1}^*) e^i(k), \quad i = 1, 2, \dots \quad (5)$$

where $e^i(k) := \sum_{j=1}^k e(\pi(j))$. We call $\theta = *$ the *star* decomposition.

We remark that either of these two decompositions may be constructed inductively by subtracting maximal multiples of $(0, 1)$ -vectors; the $(0, 1)$ -vector is of minimal support in (4), and of maximal support in (5).

We now see how simply the bound $M(B)$ on \mathcal{B} applied to decompositions (4) and (5) yields one old² and one new permenal upper bound on \mathcal{A} ,

$$U^0(A) := \prod_{i=1}^n \sum_{k=1}^n a_{i,k}^*, \quad U^M(A) := \prod_{i=1}^n \sum_{k=1}^n a_{i,k}^* \delta(k), \quad (6)$$

where for $k = 1, 2, \dots$ we set $\delta(k) := \gamma(k) - \gamma(k-1)$.

Theorem 1. *Let $U^0(A)$ and $U^M(A)$ be defined by (6). The decomposition bound based on the Minc–Brègman bound M in (1) and the decomposition $\theta = 0$ is $U^0(A)$, and that based on M and $\theta = *$ is $U^M(A)$. We also have*

$$U^0(A) \geq U^M(A) \geq \text{per}(A), \quad A \in \mathcal{A}, \quad (7)$$

$$U^M(B) = M(B), \quad B \in \mathcal{B}. \quad (8)$$

Proofs are in Section 4. The proof of Theorem 1 illustrates the basic elements used in deriving decomposition bounds on \mathcal{A} from a bound on \mathcal{B} :

² It is a bit easier to compare $\prod_i \sum_k a_{i,k}$ with other bounds when written $\prod_i \sum_k a_{i,k}^*$.

- rows of A are spanned by nonnegative combinations of $(0, 1)$ -vectors,
- $\text{per}(A)$ is a multilinear function of the rows of A , and
- the bound on \mathcal{B} is a product of functions on the rows of B .

It did not matter that the function to be bounded was the permanent, nor for that matter that it assumes only nonnegative values. However, we shall see in Theorem 2 that in order to compare a decomposition bound with other bounds, we shall require the row functions comprising M to have quite specific properties. It seems almost accidental that the Minc–Brègman bound has the properties we need.

A corollary of Theorem 1 is that the star decomposition (5) provides a 1 to 1 correspondence between the two permanental upper bounds $M(B)$ on \mathcal{B} and $U^M(A)$ on \mathcal{A} , $M \leftrightarrow U^M$; M is obtained directly from U^M by restricting its domain from \mathcal{A} to \mathcal{B} , and U^M is obtained from M through the decomposition (5).

We remark that the old and weak bound U^0 , the *trivial* bound, is easily seen to be a permanental upper bound because the monomial expansion of $U^0(A)$ has n^n nonnegative terms, $n!$ of which constitute the permanent.

1.4. Some history

The material in the next two sections will help place what follows into context. Readers familiar with this problem may wish to skip ahead.

1.4.1. Sharpening

This is our name for a general technique, for which there seems to be minimal documentation, for improving an upper bound $U(A)$ for a real-valued function $p(A)$, $A \in \mathcal{A}$, when p has invariant properties. One form of the basic idea is that we have a set \mathfrak{G} of bijections on \mathcal{A} , and a positive-valued function ψ on \mathfrak{G} , such that $p(g(A)) = \psi(g)p(A)$ holds for all $A \in \mathcal{A}$ and all $g \in \mathfrak{G}$. We called (\mathfrak{G}, ψ) a *scale-symmetry* of p on \mathcal{A} [15], and showed that if \mathfrak{G} is a group, then either U has the scale-symmetries of (\mathfrak{G}, ψ) as well, or

$$U_{\mathfrak{G}}(A) := \inf_{g \in \mathfrak{G}} U(g(A))/\psi(g)$$

is a better upper bound for p than U , with strict improvement guaranteed at certain A . (A similar argument applies to lower bounds. See [15].)

For a class of generalized matrix functions $p(A)$ which includes the permanent, examples of $g \in \mathfrak{G}$ include taking the transpose, permuting rows and/or columns, and multiplying each row and/or column by a positive scalar. When we sharpen a permanental bound $U(A)$ to obtain $U_{\mathfrak{G}}(A)$, we generally mean that \mathfrak{G} is the maximal group for which the permanent has scale-symmetries and U does not. If we let \mathfrak{S} be that subgroup of \mathfrak{G} consisting of the $n!$ row rearrangements of $A \in \mathcal{A}$, and ψ be the constant-1 function, then sharpening the bounds $U^J(A)$, $L^J(A)$ in (12) using (\mathfrak{S}, ψ)

gives the bounds $U_{\mathfrak{S}}^J(A)$ in (14) and $L_{\mathfrak{S}}^J(A)$ in (15). If we replace \mathcal{A} with \mathcal{B} , then similarly sharpening the bounds in (13) gives the bounds in (16) and (17)³.

It was also shown in [15] that if $U(A) \geq V(A) \geq p(A)$ holds on \mathcal{A} , then $U_{\mathfrak{G}}(A) \geq V_{\mathfrak{G}}(A) \geq p(A)$ holds after sharpening.

Let x denote a positive n -vector and σ denote a permutation of $1, \dots, n$. Sharpening upper bounds $U(A)$ occurs in only three ways in Section 3:

$$U_{\mathfrak{G}}([a_{i,j}]) = \inf_x x_1 \cdots x_n U([a_{i,j}/x_j]), \quad (9)$$

which applies to bounds such as $U = U^0$ and $U = U^M$ that are invariant under reordering rows;

$$U_{\mathfrak{S}}([a_{i,j}]) = \min_{\sigma} U([a_{\sigma(i),j}]), \quad (10)$$

which is a partial sharpening of bounds U that are sensitive to reordering rows (see (14) and (16)); and

$$U_{\mathfrak{R}}([a_{i,j}]) = \min_{\sigma} \inf_x x_1 \cdots x_n U([a_{\sigma(i),j}/x_j]), \quad (11)$$

which is a full sharpening of bounds U that are sensitive to reordering rows.

If our example matrices were not symmetric, then a small additional improvement could be obtained by minimizing any of the above over A and its transpose.

1.4.2. Earlier relevant permanental bounds

In 1968 [9], Minc stated that the upper and lower bounds [5]

$$U^J(A) := \prod_i \sum_{k \leq i} a_{i,k}^* \geq \text{per}(A) \geq \prod_i \sum_{k \geq n-i+1} a_{i,k}^* =: L^J(A) \quad (12)$$

were “rather surprisingly the only known nontrivial bounds for a general nonnegative matrix”. Among the trivial upper bounds, he no doubt included what we call *the* trivial (or naïve [15]) bound $U^0(A)$, the product of the row sums of A . In [5] there also appeared, as special cases of (12) applied to \mathcal{B} , the $(0, 1)$ -matrix bounds

$$J(B) := \prod_i \min(|B_i|, i) \geq \text{per}(B) \geq \prod_i \max(|B_i| - n + i, 0). \quad (13)$$

The four *Jurkat–Ryser* bounds (12) and (13) were improved by Minc [9] through minimizing the upper bounds, and maximizing the lower bounds, over permutations of the rows of the matrix in question:

$$U_{\mathfrak{S}}^J(A) := \min_{\sigma} \prod_i \sum_{k \leq i} a_{\sigma(i),k}^* \geq \text{per}(A), \quad (14)$$

$$\text{per}(A) \geq \max_{\sigma} \prod_i \sum_{k \geq n-i+1} a_{\sigma(i),k}^* =: L_{\mathfrak{S}}^J(A) \quad (15)$$

³ If p has the scale-symmetries of (\mathfrak{G}, ψ) on \mathcal{A} , then we can always sharpen a bound $U' \geq p$ which holds on a subset S of \mathcal{A} , by using those scale-symmetries for which $g \in \mathfrak{G}$ is a bijection on S as well as on \mathcal{A} .

for nonnegative matrices A , and their special cases

$$J_{\mathfrak{S}}(B) := \min_{\sigma} \prod_i \min(|B_{\sigma(i)}|, i) \geq \text{per}(B), \quad (16)$$

$$\text{per}(B) \geq \max_{\sigma} \prod_i \max(|B_{\sigma(i)}| - n + i, 0) \quad (17)$$

for $(0, 1)$ -matrices B . (The improvements (14)–(17) are special cases of sharpening; unlike the permanent, their unsharpened versions are not invariant under reordering rows.) Then Minc [10] showed that the same permutation optimizes both (16) and (17); namely, choose σ so $|B_{\sigma(i)}| = |B_i^*|$ is nonincreasing in i .

The Minc–Brègman upper bound $M(B)$ in (1), proposed in [8], remained an open question for ten years until it was resolved in [1], and is generally, but not uniformly, stronger than $J(B)$ in (13).

A succinct proof of (1) appeared later in a paper by Schrijver [13], in which he sketched a parallel proof for matrices $A \in \mathcal{A}$, and so obtained what is surely the first upper bound on \mathcal{A} which agrees with M on \mathcal{B} . Schrijver’s bound is however relatively intractable in that it appears to be more difficult to compute than the permanent itself. In the same paper he also sketched Brouwer’s decomposition idea, and gave as a consequence a second, “better”, bound, the *Brouwer–Schrijver* bound⁴ [13]

$$\prod_i \sum_k (a_{i,1}^* - a_{i,k+1}^*) \gamma(k) \geq \text{per}(A). \quad (18)$$

In the same year, Minc proved⁵ that the lower bound (17) is best possible.⁶

There followed a 20-odd year period of inactivity on upper bounds, until the first four tractable permanent upper bounds appeared [14,15] which agree with the Minc–Brègman bound M on \mathcal{B} ; they also satisfy

$$U^1(A) \geq U^2(A) \geq U^3(A) \geq U^4(A) \geq \text{per}(A), \quad A \in \mathcal{A}.$$

It was argued that these four bounds were the strongest permanent upper bounds (without sharpening) to date. The best of these, $U^4(A)$, obtained by Lagrange interpolation on the nonnegative unit cube then extension to \mathcal{A} , is given by

⁴ In examples this bound is quite poor. Because it is, in fact, sometimes worse than the trivial bound, by (22) in Theorem 2 it cannot be a consequence of decomposition as claimed. A (25 year old!) misprint is inferred; once detected the misprint is easily identified (compare (18) with (30)), and the corrected bound is $U^M(A)$, which we call the “lost” bound because it was clearly intended to appear in [13], but seems to not yet have appeared correctly in print. It is perhaps due to this unfortunate misprint that so little notice seems to have been taken of decomposition and the bounds so obtained.

⁵ Without quite saying so [11].

⁶ We note that the Jurkat–Ryser lower bound (13), being frequently 0, is often less useful than the corresponding upper bound, and the same can be said of the improved *Jurkat–Ryser–Minc* bounds (16) and (17). Yet (17) is the best possible among all lower bounds which are functions of the row sums $\{|B_i|\}$ of the matrix $B \in \mathcal{B}$. By easy examples for $n = 3$, neither (16) nor (1) is best possible.

$$U^4(A) = \prod_i t(A_i) \sum_{\ell=1}^n c_\ell E_\ell(t(A_i)^{-1} A_i), \quad A \in \mathcal{A}, \quad (19)$$

where E_ℓ is the ℓ th elementary symmetric function, $t(A_i)$ is the largest entry in the vector A_i , and

$$c_\ell := \sum_{1 \leq k \leq \ell} (-1)^{\ell-k} \binom{\ell}{k} \gamma(k). \quad (20)$$

1.5. Comparing U^M with other upper bounds

We identify the best (and worst) of all decomposition bounds $U^\theta(A)$ based on the Minc–Brègman bound $u(B) = M(B)$, and show that the best is also better than our previously best unsharpened permanental upper bound $U^4(A)$ of the previous section. Let $U^0(A)$ and $U^M(A)$ be defined by (6), and θ be a generic decomposition given by (3).

Theorem 2. *The decomposition bound $U^\theta(A)$ based on M and θ is*

$$U^\theta(A) = \prod_{i=1}^n \sum_{b^i} \theta_{i,b^i} \gamma(|b^i|), \quad A \in \mathcal{A}. \quad (21)$$

Every such bound $U^\theta(A)$ satisfies

$$U^0(A) \geq U^\theta(A) \geq U^M(A), \quad A \in \mathcal{A}. \quad (22)$$

Also, $U^M(A)$ is uniformly better than $U^4(A)$ as given by (19),

$$U^4(A) \geq U^M(A) \geq \text{per}(A), \quad A \in \mathcal{A}. \quad (23)$$

So the best decomposition bound based on $M(B)$ is $U^M(A)$, given by the star decomposition $\theta = *$, and the worst such bound is given by the trivial decomposition $\theta = 0$.

We note that the only properties of the Minc–Brègman bound $M(B) = \prod_i \gamma(|B_i|)$ needed to obtain the comparisons of Theorem 2 were $\gamma(0) = 0$, $\gamma(1) = 1$, and the fact that $\gamma(k) - \gamma(k-1)$ is nonincreasing for $k \geq 1$.

2. Decomposition bounds in general

Thus far our decomposition bounds have all been based on the Minc–Brègman bound $M(B)$, and the best of these bounds, $U^M(A)$, after sharpening will remain our strongest bound when all is said and done. However, the following generalization of Theorems 1 and 2 should be of more than passing theoretical interest.

In this section only, row indices i range from 1 to m . We wish to bound a real-valued row-multilinear function $p(A)$ defined on the (entrywise) nonnegative m by n matrices \mathcal{A} , but are only provided with an upper bound $u(B)$ for $p(B)$ on the $(0, 1)$ -matrices $\mathcal{B} \subset \mathcal{A}$, of the form

$$u((B_i)) = \prod_i u_i(|B_i|) \geq p((B_i)), \quad (B_i) \in \mathcal{B}. \quad (24)$$

Let a generic row decomposition θ of $A \in \mathcal{A}$ be given by (3), the trivial decomposition $\theta = 0$ be given by (4), and the star decomposition $\theta = *$ be given by (5). For each i set $\delta_i(k) := u_i(k) - u_i(k-1)$ for $k = 1, 2, \dots$. The bound u in (24) is said to be *feasible* if for each i we have $u_i(0) = 0$, $u_i(1) = 1$, and $\delta_i(k) \geq \delta_i(k+1)$ for $k = 1, 2, \dots$.

With a few minor changes of notation, we have already proved parts (1°)–(5°) of the following, a restatement of Theorems 1 and 2 in the general setting.

Theorem 3. *Statements (1°)–(6°) hold for bounds u given by (24).*

(1°) *The trivial decomposition bound based on u and $\theta = 0$ (4) is*

$$U^0(A) = \prod_i \sum_{k=1}^n a_{i,k}^* \geq p(A), \quad A \in \mathcal{A}. \quad (25)$$

(2°) *The generic decomposition bound based on u and θ (3) is*

$$U^\theta(A) = \prod_i \sum_{b^i} \theta_{i,b^i} u_i(|b^i|) \geq p(A), \quad A \in \mathcal{A}. \quad (26)$$

(3°) *The star-decomposition bound based on u and $\theta = *$ (5) is*

$$U^*(A) = \prod_i \sum_{k=1}^n a_{i,k}^* \delta_i(k) \geq p(A), \quad A \in \mathcal{A}. \quad (27)$$

(4°) *For $B \in \mathcal{B}$ we have $U^*(B) = u(B)$.*

(5°) *If u is feasible, then for $A \in \mathcal{A}$ we have*

$$U^0(A) \geq U^\theta(A) \geq U^*(A), \quad (28)$$

$$U^4(A) \geq U^*(A) \geq p(A). \quad (29)$$

(6°) *If u is feasible and no $u_i(k)$ is constant in k , then the only decomposition bound U^θ which agrees with u on \mathcal{B} is the star bound $U^\theta = U^*$.*

The required changes in notation include p for per; u for M and $u_i(k)$ for $\gamma(k)$ (such as in (20)); and U^* for U^M when the star bound is based on a generic u , reserving U^M for the case $u = M$. Finally, $\delta_i(k) = u_i(k) - u_i(k-1)$ replaces $\delta(k) = \gamma(k) - \gamma(k-1)$.

A corollary of Theorem 3 is that the star decomposition provides a 1 to 1 correspondence between the two permanental upper bounds $u(B)$ on \mathcal{B} and $U^*(A)$ on \mathcal{A} , $u \leftrightarrow U^*$; u is obtained directly from U^* by restricting its domain from \mathcal{A} to \mathcal{B} , and

U^* is obtained from u through the star decomposition (5). Indeed the two bounds u and U^* are the same in the sense that

$$u(B) \geq p(B) \quad \forall B \in \mathcal{B} \quad \text{iff} \quad U^*(A) \geq p(A) \quad \forall A \in \mathcal{A}.$$

2.1. Three applications of Theorem 3

We show that the Minc–Brègman bound M , the Jurkat–Ryser bound J , and the trivial bound U^0 itself are feasible bounds on \mathcal{B} . For these specific u 's we write the resulting star bound $U^*(A)$ as $U^M(A)$, $U^J(A)$, and $U^0(A)$ respectively.

If $u_i(k) = \gamma(k)$, then from (1) and (24) we see that $u(B)$ is the Minc–Brègman bound $M(B)$. The feasibility of M is a consequence of the concavity of $\Gamma(x+1)^{1/x}$, a result of Sandor [12].

If $u_i(k) = \min(k, i)$, then from (13) and (24) we see that $u(B)$ is the Jurkat–Ryser bound $J(B)$. The feasibility of J is immediate since $u_i(k) - u_i(k-1)$ equals 1 if $k \leq i$ and 0 if $k > i$, so is nonincreasing. For the special case $u = J$ we recognize $U^J(A)$ in (12) as the star bound (27).

If $u_i(k) = k$, then directly from (24) we see that $u(B)$ is the trivial bound $U^0(B)$. The feasibility of U^0 is trivially verified. For the special case $u = U^0$, $U^*(A)$ is the trivial bound $U^0(A)$, so *each* bound $U^\theta(A)$ is the trivial bound.

The three star-decomposition bounds for $u = M$, $u = J$, and $u = U^0$ may be written, recalling $\delta(k) = \gamma(k) - \gamma(k-1)$,

$$\begin{aligned} u = M : \quad U^*(A) = U^M(A) &:= \prod_i \sum_{k=1}^n a_{i,k}^* \delta(k), \\ u = J : \quad U^*(A) = U^J(A) &:= \prod_i \sum_{k=1}^i a_{i,k}^*, \\ u = U^0 : \quad U^*(A) = U^0(A) &:= \prod_i \sum_{k=1}^n a_{i,k}^*. \end{aligned}$$

It seems a bit odd that a bound so weak as the trivial bound U^0 would occur naturally in our development, yet it does so in *three* ways: as a feasible bound on \mathcal{B} , as a decomposition bound based on any bound of the form (24) paired with the trivial decomposition $\theta = 0$, and as a decomposition bound based on $u(B) = U^0(B)$ paired with any decomposition θ .

3. Examples

To illustrate the effectiveness of the bounds in question, we give two examples where a formula for the permanent is known: one matrix $B \in \mathcal{B}$, and one matrix $A \in \mathcal{A}$ not in \mathcal{B} .

For integers $n \geq k \geq 1$ consider the matrix $B(k) \in \mathcal{B}$ having i, j entry 1 if and only if $i + j \leq n + k$. An easy induction shows that $\text{per}(B(k)) = k!k^{n-k}$. As k increases to n , $B(k)$ approaches the constant-1 matrix, so from the case of equality $M(B) = \text{per}(B)$, one might expect the “fit ratio” for M at $B(k)$, $M(B(k))/\text{per}(B(k))$, to grow as k decreases from n to 1. In [15] we found that the fit ratio grows from 1 to $\prod_{j=1}^n j^{1/j}$ as k decreases from n to 1. We give the ‘hard’ case $k = 2$, $n = 36$ and set $B := B(2)$.

Another matrix with known permanent is obtained from $B(1)$ above by replacing every 0 with z , $z \geq 0$. That is, $a_{i,j}(z) = z$ for $i + j > n + 1$, and 1 otherwise. Let $E = E(z)$ be the matrix obtained from $A(z)$ by reversing the order of its rows. So E has 1’s on and below the main diagonal, and z ’s above, and has the same permanent as $A(z)$.

A formula for $\text{per}(A(z)) = \text{per}(E)$ is

$$\text{per}(E) = \sum_{k=0}^{n-1} \left\langle \begin{matrix} n \\ k \end{matrix} \right\rangle z^k,$$

which follows from the definition (see [6]) of the k th Eulerian number $\left\langle \begin{matrix} n \\ k \end{matrix} \right\rangle$, being equivalent to [6, p. 46] the number of permutations σ on $1, \dots, n$ having exactly k values i for which $i < \sigma(i)$ ($1 \leq i < n$), $0 \leq k < n$. Notice that $\text{per}(A(z))$ ranges from 1 at $z = 0$ to $n!$ at $z = 1$. We give the case $z = 1/6$, $n = 36$, and set $A := A(1/6)$.

For each upper bound in Table 1 we give a reference “Ref” for its definition. Recall for a bound U that the subscript \mathfrak{G} means U is sharpened by optimizing over column rescalings (see (9)); the subscript \mathfrak{H} means U is sharpened by optimizing over row reorderings (see (10)); and the subscript \mathfrak{R} means U is sharpened by optimizing over both (see (11)). For these bounds and matrices, none of the other scale-symmetries in [15] apply.

For lower bounds g in Table 2 we have included the (usually) strong permanental lower bound $L(A)$, which we write as $L(A) := n!n^{-n}U_{\mathfrak{G}}^0(A) \leq \text{per}(A)$; the inequality $\text{per}(A) \geq L(A)$ is equivalent to the Egoritsjev–Falikman theorem [2,3] (formerly the van der Waerden conjecture) and has been used in applications since the 1980s (e.g., [4]), yet may not have appeared formally until [7]. We note that L has all the scale-symmetries of the permanent, so is not a candidate for improvement by sharpening.

Four things to notice in the B example are that M is not listed as it agrees with U^1 , U^4 and U^M (just as J agrees with U^J); that B is in the “best possible” realm of L^J ; that B already has the optimal row ordering for sharpening by \mathfrak{H} [10], so that $U_{\mathfrak{H}}^J(B) = U^J(B)$; and that sharpening performs very well here, especially $U_{\mathfrak{G}}^0$ and $U_{\mathfrak{R}}^J$, which are far better than M itself. One consequence of this last observation is that since $U_{\mathfrak{G}}^0/L = n^n/n!$ depends on n only, the bound L , which is not amenable to sharpening, performs much better on the A example than on the B example.

Table 1
Comparing upper bounds with $\text{per}(B)$, $\text{per}(A)$

Ref	f	$f(B)$	$f(B)/\text{per}(B)$	$f(A)$	$f(A)/\text{per}(A)$
(25)	U^0	1.3392×10^{43}	3.8975×10^{32}	2.3981×10^{46}	2.15×10^{16}
(25), (9)	$U_{\mathbb{G}}^0$	1.1806×10^{21}	3.4359×10^{10}	2.7910×10^{44}	2.50×10^{14}
(12)	U^J	7.7882×10^{32}	2.2666×10^{22}	1.6918×10^{35}	151607
(12), (10)	$U_{\mathbb{S}}^J$	7.7882×10^{32}	2.2666×10^{22}	4.7176×10^{34}	42276
(12), (11)	$U_{\mathbb{R}}^J$	1.3054×10^{20}	3.7994×10^9	8.6298×10^{33}	7733
[14]	U^1	1.5973×10^{30}	4.6487×10^{19}	1.2331×10^{33}	1105
(19)	U^4	1.5973×10^{30}	4.6487×10^{19}	6.4951×10^{32}	582
(6)	U^M	1.5973×10^{30}	4.6487×10^{19}	5.7669×10^{32}	517
[15], (9)	$U_{\mathbb{G}}^1$	3.0126×10^{15}	87677	1.9316×10^{31}	17.3
(19), (9)	$U_{\mathbb{G}}^4$	6.6543×10^{14}	19366	8.7954×10^{30}	7.88
(6), (9)	$U_{\mathbb{G}}^M$	5.6289×10^{14}	16382	7.2793×10^{30}	6.52
	Per	3.4359×10^{10}	1.0	1.1159×10^{30}	1.0

Table 2
Comparing lower bounds

Ref	g	$\text{per}(B)/g(B)$	$\text{per}(A)/g(A)$
Above	L	8323	1.1434
(12)	$L_{\mathbb{G}}^J$	1.0	182
(15)	$L_{\mathbb{S}}^J$	1.0	182
(12)	L^J	1.0	4.1288×10^{10}

Bounds have been ranked according to their value on A , which example seems overall to be the more typical of the two.

4. Proofs

Proof of Theorem 1. In (5), set $\Delta_{i,k} := a_{i,k}^* - a_{i,k+1}^*$ and replace k with k_i . Noting $|e^i(k)| = k$, the row-multilinearity of the permanent of $A = (A_i)$ as expressed by decomposition $\theta = *$, given by (5), yields by definition

$$\text{per}(A) = \sum_{k_1} \cdots \sum_{k_n} \Delta_{i,k_1} \cdots \Delta_{n,k_n} \text{per}((e^i(k_i))),$$

where each index k_i runs from 1 to n for every i . As each Δ product is nonnegative, replacing each term $\text{per}((e^i(k_i)))$ with its upper bound $M((e^i(k_i))) = \prod_i \gamma(k_i)$ yields an expression which bounds $\text{per}(A)$

$$\begin{aligned} \text{per}(A) &\leq \sum_{k_1} \cdots \sum_{k_n} \Delta_{i,k_1} \cdots \Delta_{n,k_n} M((e^i(k_i))) \\ &= \sum_{k_1} \cdots \sum_{k_n} \prod_i \Delta_{i,k_i} \gamma(k_i), \end{aligned}$$

which, by a standard formula for factoring such sums, becomes

$$= \prod_i \sum_{k_i} \Delta_{i,k_i} \gamma(k_i), \quad (30)$$

and by reversing the collapsing sums we get

$$= \prod_i \sum_k a_{i,k}^* \delta(k) =: U^M(A). \quad (31)$$

This proves the right-hand inequality in (7).

Starting instead with $\theta = 0$, given by (4), and noting $|e(k)| = 1$, we get

$$\begin{aligned} \text{per}(A) &= \sum_{k_1} \cdots \sum_{k_n} a_{i,k_1} \cdots a_{n,k_n} \text{per}((e(k_i))) \\ &\leq \sum_{k_1} \cdots \sum_{k_n} a_{i,k_1} \cdots a_{n,k_n} M((e(k_i))) \\ &= \sum_{k_1} \cdots \sum_{k_n} \prod_i a_{i,k_i} \gamma(1), \\ &= \prod_i \sum_{k_i} a_{i,k_i} \\ &= \prod_i \sum_k a_{i,k}^* =: U^0(A). \end{aligned} \quad (32)$$

Comparing (31) with (32) we see that $U^0(A) \geq U^M(A)$ is a consequence of $\delta(k) \leq 1$, which in turn is a consequence of the concavity [12] of $\Gamma(x+1)^{1/x}$. This proves the left-hand inequality in (7).

The fact that U^M and M agree on \mathcal{B} is a direct consequence of every row b of B^* , $B \in \mathcal{B}$, having $|b|$ ones followed by $n - |b|$ zeros; for such a row, from (31) we get $\sum_k b_k \delta(k) = \sum_{k=1}^{|b|} \delta(k) = \gamma(|b|)$. This proves (8). \square

Proof of Theorem 2. Given the bound M for per on \mathcal{B} satisfying $M((b_i)) = \prod_i \gamma(|b^i|)$, the derivation of (21) from (3) parallels the derivations of U^M and U^0 in Theorem 1. The row-multilinearity of the permanent applied to the row decomposition θ in (3) gives, by definition,

$$\begin{aligned}
\text{per}(A) &= \sum_{b^1, \dots, b^n} \text{per}((b^i)) \prod_{i=1}^n \theta_{i, b^i} \\
&\leq \sum_{b^1, \dots, b^n} M((b^i)) \prod_{i=1}^n \theta_{i, b^i} \\
&= \sum_{b^1, \dots, b^n} \prod_{i=1}^n \theta_{i, b^i} \gamma(|b^i|) \\
&= \prod_{i=1}^n \sum_{b^i} \theta_{i, b^i} \gamma(|b^i|) \\
&=: \prod_{i=1}^n T_i^\theta(b^i) =: U^\theta((b^i)).
\end{aligned}$$

This proves (21); an interpretation is that every decomposition bound $U^\theta(A)$ may be written as a product of row- i factors $T_i^\theta(A_i)$. We abbreviate this notation by saying if r is the i th row of A having decomposition $r = \sum_b \theta_b b$, then

$$T_i^\theta(r) = \sum_b \theta_b \gamma(|b|) = \sum_b \theta_b \sum_{k=1}^{|b|} \delta(k). \quad (33)$$

Our only method for comparing bounds is through comparing their corresponding row factors. And to compare, for example, the i th row factors $T_i^M(r)$ and $T_i^\theta(r)$ of U^M and U^θ respectively, we use (2) to express $T_i^M(r)$ in terms of θ , as follows.

Again let π be a permutation for which $(r_{\pi(j)}) = r^*$ has nonincreasing entries. It follows from (2) that $r_j = \sum_b \theta_b b_j$ and so $r_j^* = r_{\pi(j)} = \sum_b \theta_b b_{\pi(j)}$ hold, so that from (6) we get

$$\begin{aligned}
T_i^M(r) &= \sum_k r_k^* \delta(k) \\
&= \sum_k \delta(k) \sum_b \theta_b b_{\pi(k)} \\
&= \sum_b \theta_b \sum_k \delta(k) b_{\pi(k)} = \sum_b \theta_b \sum_{k: b_{\pi(k)}=1} \delta(k).
\end{aligned} \quad (34)$$

From (6) and (34) we also infer that

$$T_i^0(r) = \sum_b \theta_b \sum_k b_{\pi(k)} = \sum_b \theta_b |b| \quad (35)$$

are the corresponding expressions for $T_i^0(r)$ in terms of θ .

From (35), (33) and (34) we obtain $T_i^0(r) - T_i^\theta(r) = \sum_b \theta_b C(b)$ and $T_i^\theta(r) - T_i^M(r) = \sum_b \theta_b D(b)$ where

$$C(b) = |b| - \sum_{k=1}^{|b|} \delta(k) \quad \text{and} \quad D(b) = \sum_{k=1}^{|b|} \delta(k) - \sum_{k: b_{\pi(k)}=1} \delta(k).$$

We have already noted that the δ 's are nonincreasing, and as $\delta(1) = 1$ the nonnegativity of both $C(b)$ and $D(b)$ follow, proving (22).

The permanental upper bound U^4 on \mathcal{A} was obtained in [15] by showing that

$$f^4((A_i)) := \prod_{i=1}^n \sum_{\ell=1}^n c_\ell E_\ell(A_i) \geq \text{per}((A_i))$$

holds on the unit cube $\mathcal{C} := \{A : 0 \leq a_{ij} \leq 1\}$, then mapping $A \in \mathcal{A}$ into \mathcal{C} in an optimal way, and using (row) scale-symmetries of the permanent. Since U^M has the same row-scaling property as the permanent, it suffices to prove that $f^4(A) \geq U^M(A)$ holds for every A in \mathcal{C} . Again it suffices to show this inequality holds between each row- i factor; namely that, as before, $T_i^4(r) := \sum_{\ell=1}^n c_\ell E_\ell(r) \geq T_i^M(r)$ holds for $r \in \mathcal{C}$. That is, we wish to show that the difference

$$c := \sum_{\ell=1}^n c_\ell E_\ell(r) - \sum_{k=1}^n r_k^* \delta_i(k)$$

is nonnegative on \mathcal{C} .

If c were an affine function of each variable r_1, r_2, \dots, r_n , then by Lagrange interpolation it would suffice to show $c \geq 0$ holds if $r = b$ is any $(0, 1)$ -vector. The problem is that T_i^* is a convex function of each r_j , since the coefficient of r_j depends on its rank order of among all the row entries. However, we notice that $T_i^4(r)$ is independent of the order of the row entries, so that

$$c = \sum_{\ell=1}^n c_\ell E_\ell(r^*) - \sum_{k=1}^n r_k^* \delta(k)$$

holds, and to prove $c \geq 0$ holds for *all* r it suffices in this last equation to replace r_j^* with r_j . As the form of c becomes affine in each variable, it suffices to prove that

$$\sum_{\ell=1}^n c_\ell E_\ell(b) \geq \sum_{k=1}^n b_k \delta(k)$$

holds for every $(0, 1)$ -vector b . The right-hand side is $\sum_{k: b_k=1} \delta(k)$.

If $|b| = d$, the left-hand side is

$$\sum_{\ell=1}^d c_\ell \binom{d}{\ell} = \sum_{\ell=1}^d \sum_{1 \leq k \leq \ell} (-1)^{\ell-k} \binom{\ell}{k} \gamma(k) \binom{d}{\ell}$$

$$\begin{aligned}
&= \sum_{k=1}^d \gamma(k) \sum_{\ell=k}^d (-1)^{\ell-k} \binom{\ell}{k} \binom{d}{\ell} \\
&= \gamma(d) = \sum_{k=1}^d \delta(k)
\end{aligned}$$

since $\sum_{\ell=k}^d (-1)^{\ell-k} \binom{\ell}{k} \binom{d}{\ell} = \delta_{k,d}$. Thus the left-hand side is the sum of the d largest δ 's, while the right-hand side is the sum of d distinct δ 's, and $c \geq 0$ follows, proving (23). \square

Proof of Theorem 3. After the indicated changes of notation, the proofs of (25), (27), (26), part (4°), (28), and (29), appear respectively as (6) in Theorem 1, (21) in Theorem 2, (8) in Theorem 1, (22) in Theorem 2, and (23) in Theorem 2.

To prove part (6°) it suffices to consider matrices $B \in \mathcal{B}$ without zero rows, for which perforce every row factor $T_i(B_i)$, and so every bound, is positive. The proof requires careful study of the case of equality in $T_i^\theta(r) - T_i^*(r) \geq 0$ (r a $(0, 1)$ -vector).

We first note that the star decomposition of a $(0, 1)$ -vector B_i always takes the form $B_i = B_i$. Thus we are to show that if T_i^θ agrees with u_i , and so with T_i^* , on the set of $(0, 1)$ -vectors, and the decomposition θ on row i is not $B_i = B_i$, then $\delta_i(k)$ is constant in k .

For the decomposition θ set

$$s_k := \sum_{b: |b| \geq k} \theta_b.$$

The sequence s_k is nonincreasing, with $s_1 > 0$ and the largest k for which $s_k > 0$ being $k^* = \max_b |b|$. If $|B_i| = d$, it follows from $b_{ij} = \sum_b \theta_b b_j$ that $\sum_{k=1}^d s_k = d$. Reversing the order of summation, we find $T_i^\theta = \sum_{k=1}^{k^*} s_k \delta_i(k)$ while $T_i^* = \sum_{k=1}^d \delta_i(k)$. We have $d \geq k^*$ and

$$T_i^\theta - T_i^* = \sum_{k=1}^d (s_k - 1) \delta_i(k) \geq 0. \quad (36)$$

Notice that $k^* = d$ and $s_1 = \dots = s_d$ is not possible, otherwise each b would have $|b| = d$, then the “distinct b 's” property would require θ have a single summand $B_i = B_i$, so that $\theta = *$. It follows that $s_1 > 1$ and $s_d < 1$.

Suppose $s_k = 1 + \alpha_k$ with $\alpha_k > 0$ for $k \leq \ell$, $s_k = 1 - \beta_k$ with $\beta_k > 0$ for $k \geq m$, and $s_k = 1$ in between. Equality in (36) is the same as

$$\sum_{k \leq \ell} \alpha_k \delta_i(k) = \sum_{k \geq m} \beta_k \delta_i(k),$$

where $\sum_{k \leq \ell} \alpha_k = \sum_{k \geq m} \beta_k$ also hold. Thus a positive, convex combination of $\delta_i(1), \dots, \delta_i(\ell)$ equals a positive, convex combination of $\delta_i(m), \dots, \delta_i(d)$, which can only happen if each such $\delta_i(\cdot)$ takes the same value, so $\delta_i(1) = \dots = \delta_i(d)$. As B_i is arbitrary, this must hold for $d = n$, and part (6°) is proved. \square

Acknowledgments

The author would like to thank the referees for their thoughtful suggestions which greatly improved the exposition.

References

- [1] L.M. Brègman, Some properties of nonnegative matrices and their permanents, *Soviet Math. Dokl.* 14 (4) (1973) 945–949.
- [2] G.P. Egoritsjev, Solution of van der Waerden's permanent conjecture, (Russian) translation in: *Adv. Math.* 42 (1981) 299–305, preprint 1 3M of the Kirenski Institute of Physics, Krasnojarsk.
- [3] D.I. Falikman, A proof of the van der Waerden's conjecture on the permanent of a doubly stochastic matrix, *Mat. Zametki* 29 (1981) 931–938 (in Russian).
- [4] R. Grone, A consequence of the van der Waerden conjecture, personal communication, 1989.
- [5] W.B. Jurkat, H.J. Ryser, Matrix factorizations of determinants and permanents, *J. Algebra* 3 (1966) 1–27.
- [6] D. Knuth, *Sorting and Searching, The Art of Computer Programming*, 2nd ed., Addison–Wesley, Menlo Park, 1998.
- [7] N. Linial, A. Samorodnitsky, A. Wigderson, A deterministic strongly polynomial algorithm for matrix scaling and approximate permanents, in: *SPOC 98*, 1998, pp. 644–652.
- [8] H. Minc, Upper bounds for permanents of $(0, 1)$ -matrices, *Bull. Amer. Math. Soc.* 69 (1963) 789–791.
- [9] H. Minc, Bounds for permanents of nonnegative matrices, *Proc. Edinburgh Math. Soc.* 16 (1968) 233–237.
- [10] H. Minc, Rearrangements, *Trans. Amer. Math. Soc.* 159 (1971) 233–237.
- [11] H. Minc, Permanents, in: *Encyclopedia of Mathematics and its Applications*, vol. 6, Addison–Wesley, Reading MA, 1978.
- [12] J. Sandor, Sur la fonction gamma, *Publ. Cent. Rech. Math. Pures, Sér. I* 21 (1989) 4–7.
- [13] A. Schrijver, A short proof of Minc's conjecture, *J. Combin. Theory Ser. A* 25 (1978) 80–83.
- [14] G.W. Soules, Extending the Minc–Brègman upper bound for the permanent, *Linear and Multilinear Algebra* 47 (1) (2000) 77–92.
- [15] G.W. Soules, New permanental upper bounds for nonnegative matrices, *Linear and Multilinear Algebra* 51 (4) (2003) 319–337.